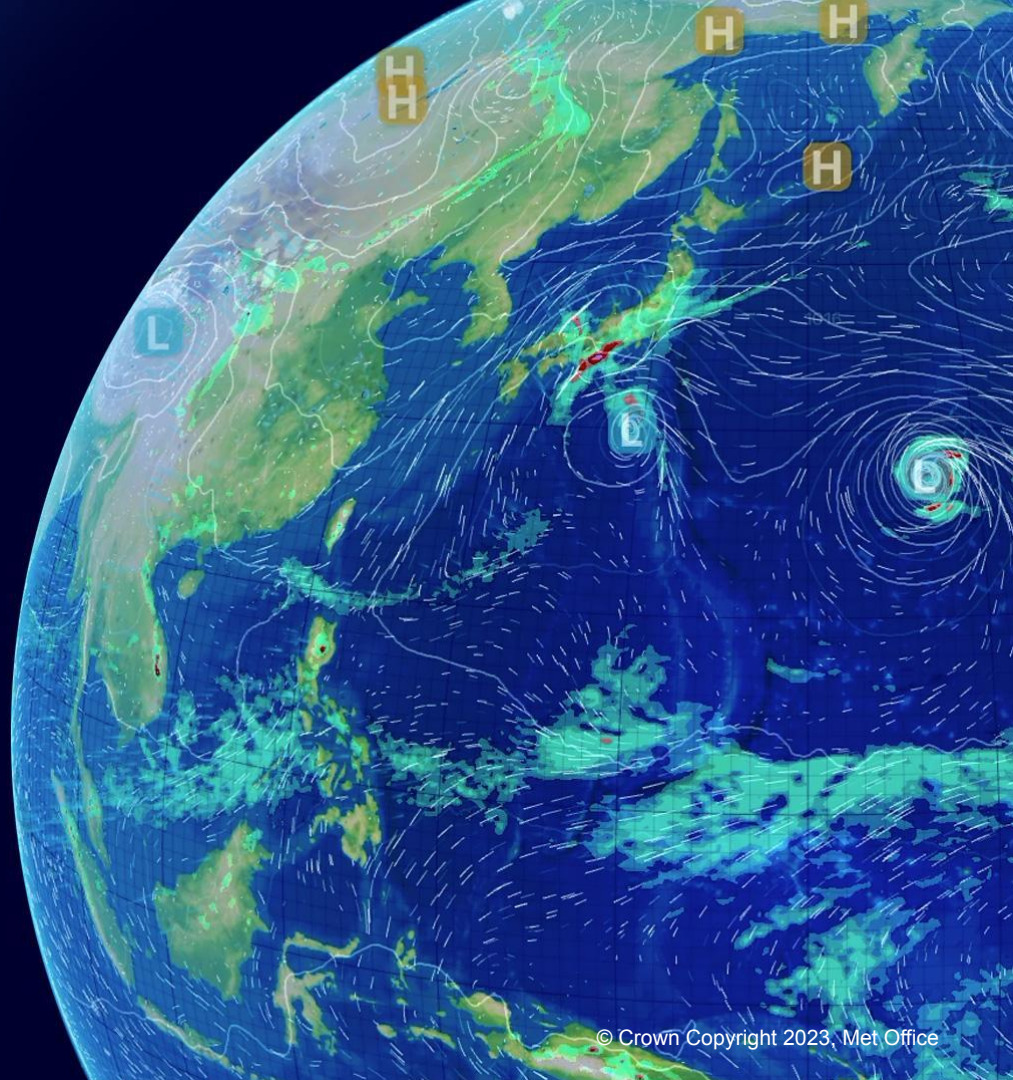


Ensemble Size: Determining the Optimal Number of Members for NWP in Research and Development

Charlotte Wynn

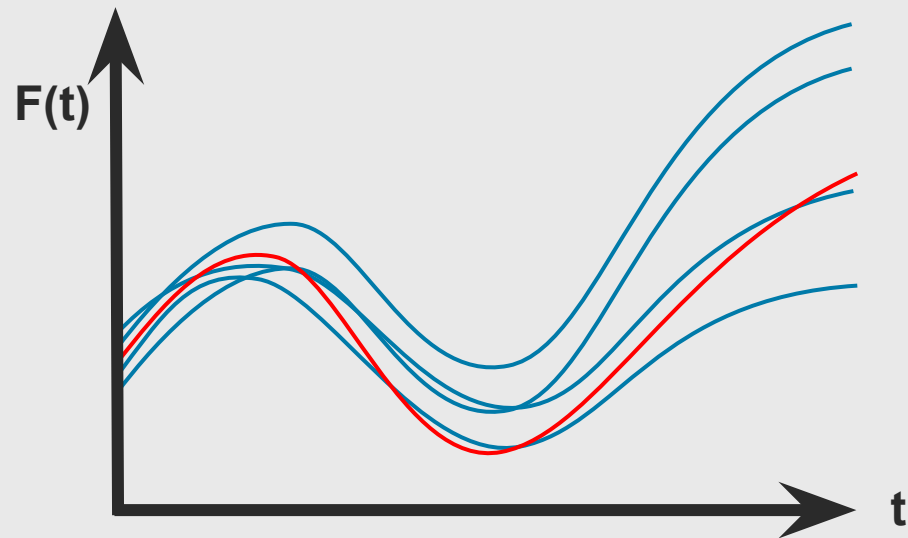


Talk outline

- Ensembles in Global NWP R&D
- Investigating ensemble size and its impact on forecast skill
- Evaluating the computational costs of ensembles in R&D
- Weighing the impacts of ensemble size against those of resolution

Ensembles in R&D

- Ensemble = 1 control + a number of perturbed members
- Perturbations in initial conditions
 - spread in forecasts
- Ensembles are computationally expensive!
- More members = higher computational cost
- What is the balance?



CRPS

•

$$\text{CRPS} = \int_{-\infty}^{\infty} [F(y) - F_o(y)]^2 dy$$

Where...

$$F_o(y) = \begin{cases} 0, & y < \text{observed value} \\ 1, & y \geq \text{observed value} \end{cases}$$

- Rewards forecasts with probability concentration around the observation
- Penalises too little/too much spread

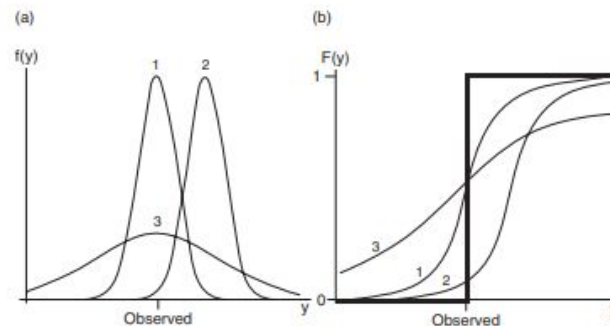


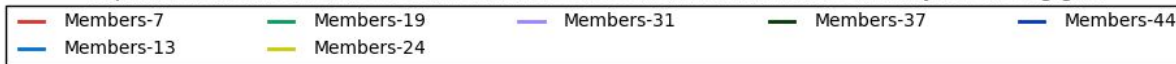
FIGURE 7.13 Schematic illustration of the continuous ranked probability score. Three forecast PDFs are shown in relation to the observed outcome in (a). The corresponding CDFs are shown in (b), together with the step-function CDF for the observation $F_o(y)$ (heavy line). Distribution 1 would produce a small (good) CRPS because its CDF is the closest approximation to the step function. Distribution 2 concentrates probability away from the observation, and Distribution 3 is penalized for lack of sharpness even though it is centered on the observation.

Trial setup

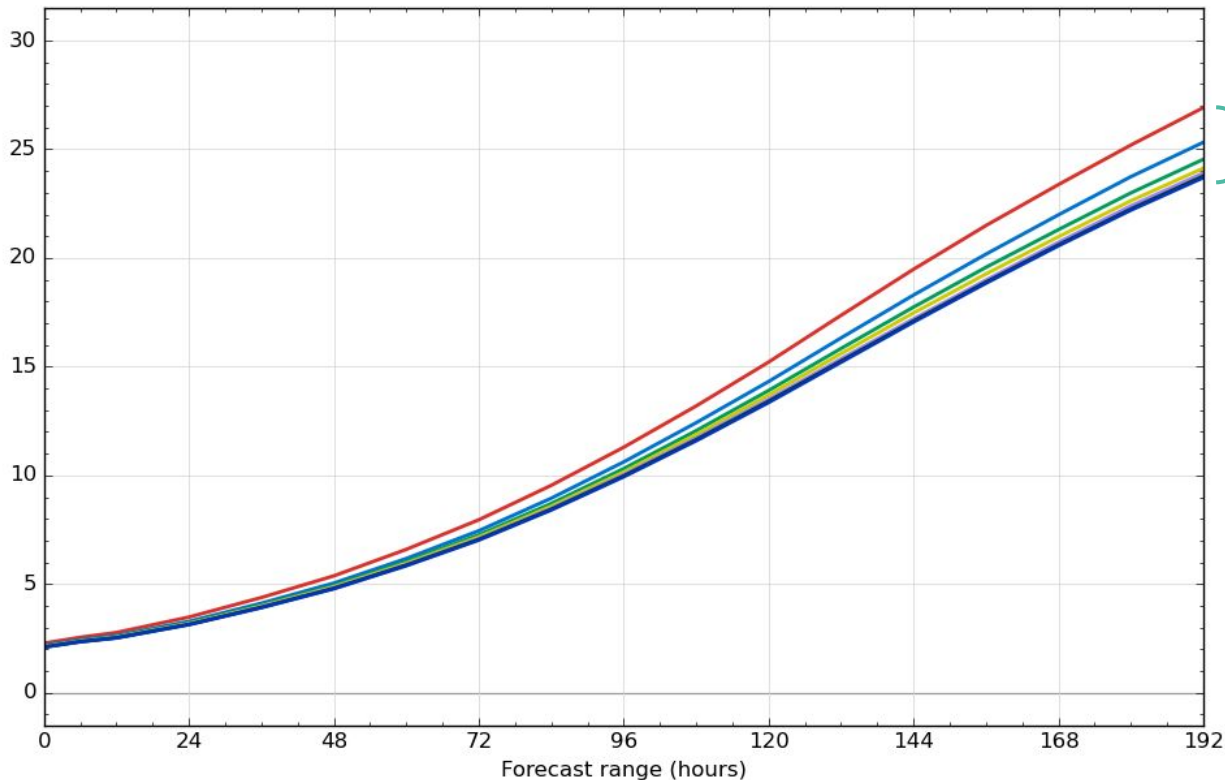
- Ensemble of 44 members (MOGREPS-G)
 - 1 control member + 43 perturbed members
- Vary number of members in verification
 - [7, 13, 19, 24, 31, 37, 44]
- Trial at different resolutions and time periods

Trial ID	Period	Resolution
u-cu111	Winter	High (~20km)
u-cs219	Summer	Low (~60km)
u-da394	Winter	Low (~60km)

Geopotential Height (m) @ 850hPa,
Continuous Ranked Probability Score (Ensemble FC(j) (Excluding Control)),
Northern Hemisphere (CBS area 90N-18.75N),
Equalized and Meaned between 20211201 00:00 and 20220214 12:00, Analysis, 1.5deg grid



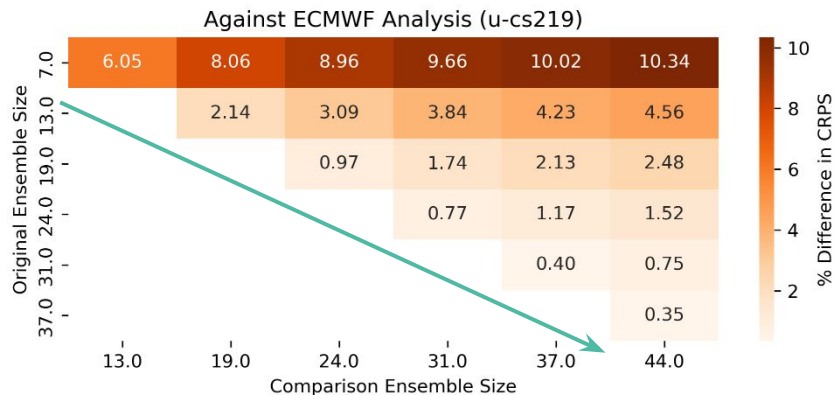
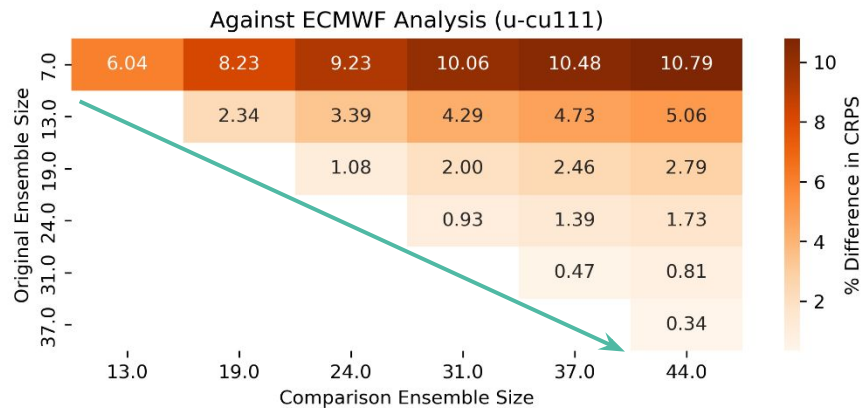
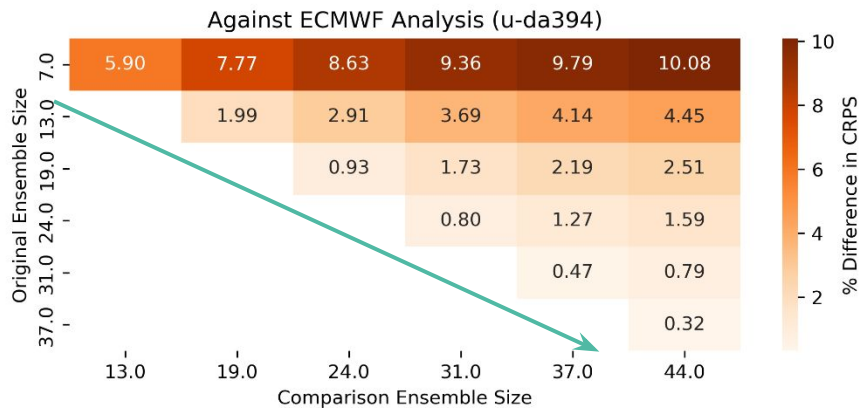
u-cu111 - Winter High



Converging towards fair CRPS

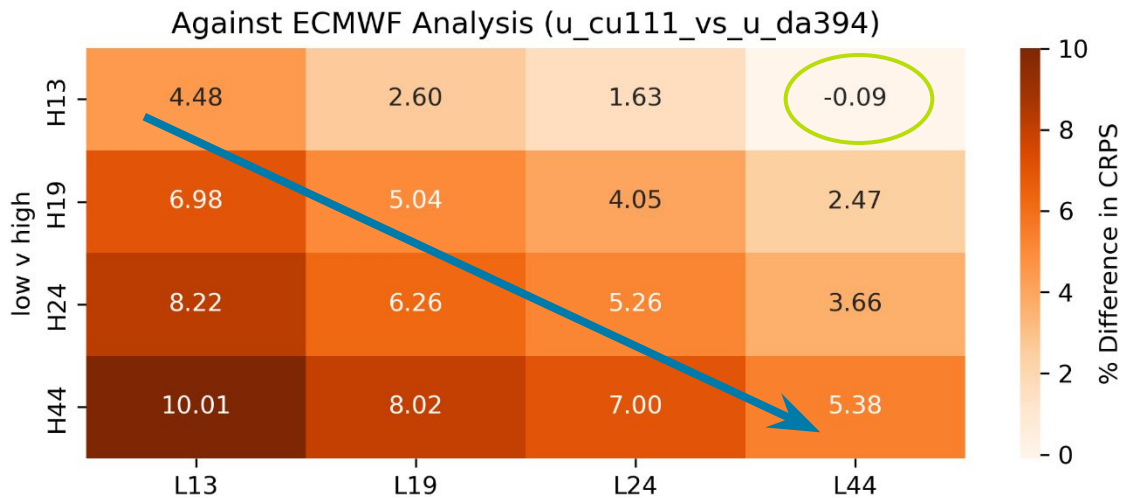


Non-linear



Comparable size increase

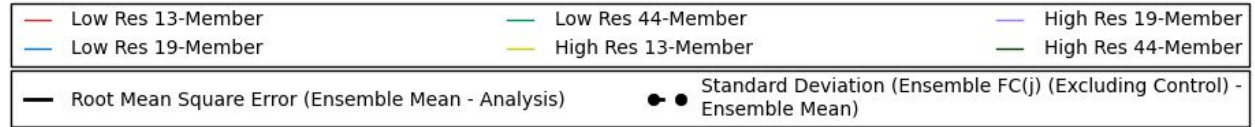
High- vs Low-
res at same size
□ ~5%
improvement
due to resolution



High-13 vs low-44:
is many low-res
better than few
high-res?

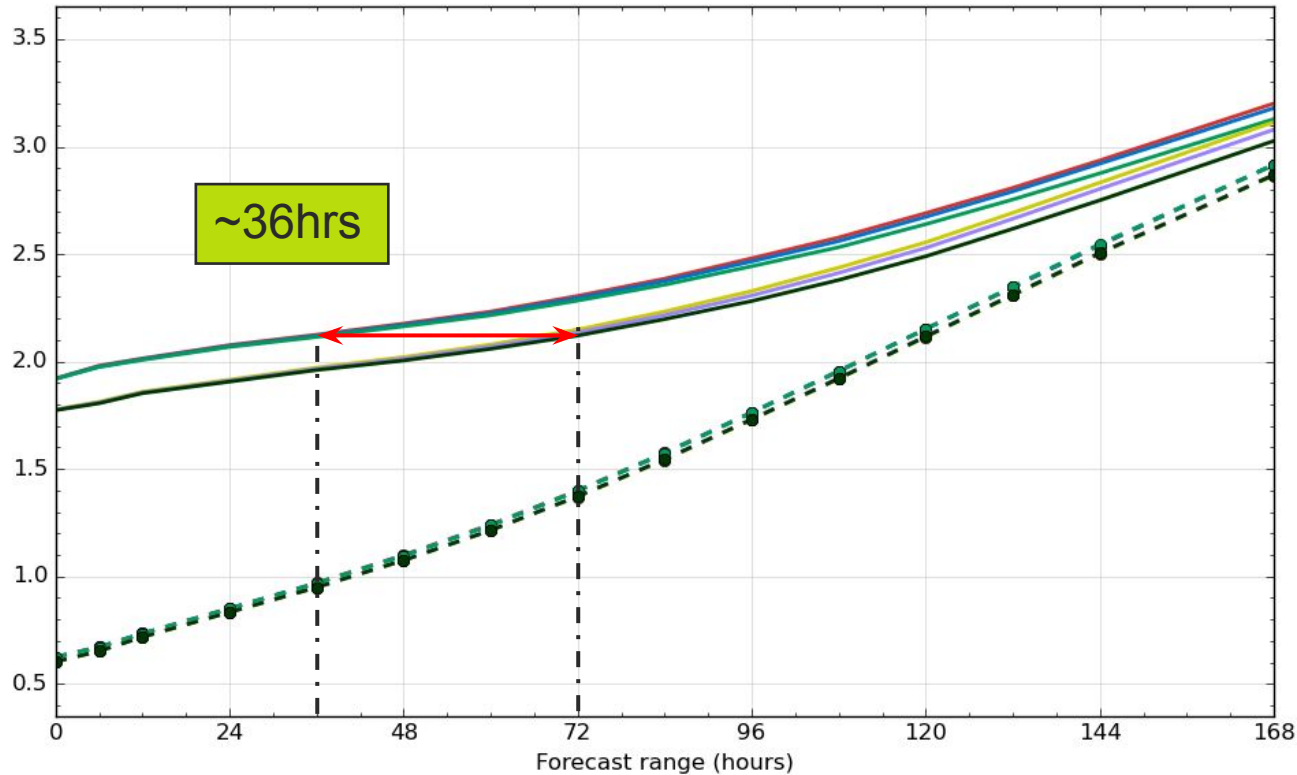
Skill vs Cost

- Comparing node hours for winter forecast cases:
 - u-da394 (low): 12 nodes * 50 min = ~ 10 Node Hours
 - u-cu111 (high): 45 nodes * 1 hr 36 min = ~ 72 Node Hours
- High-res vs low-res is more expensive by a **factor of ~7X**
- 44-mem low-res forecast is cost comparable to 6-mem high-res forecast



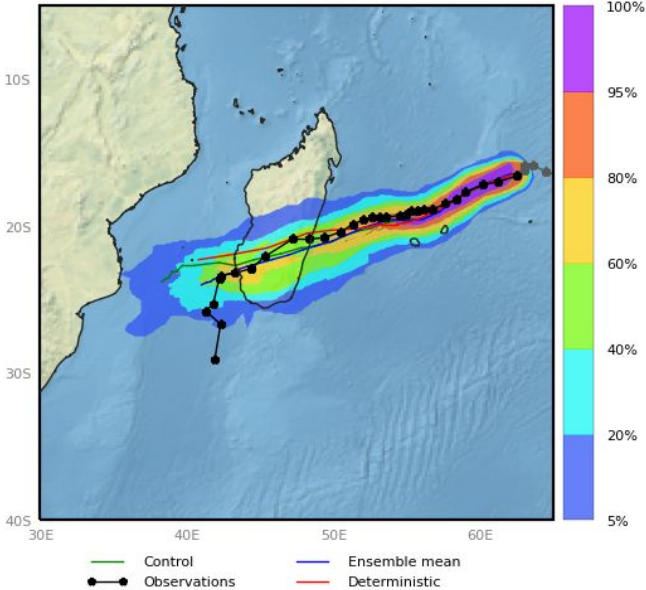
Resolution

- Significant skill difference in surface phenomena
- T+36hrs low-res has same skill as T+72hr high-res
- T+168hrs Low-44 comparable RMSE to High-13

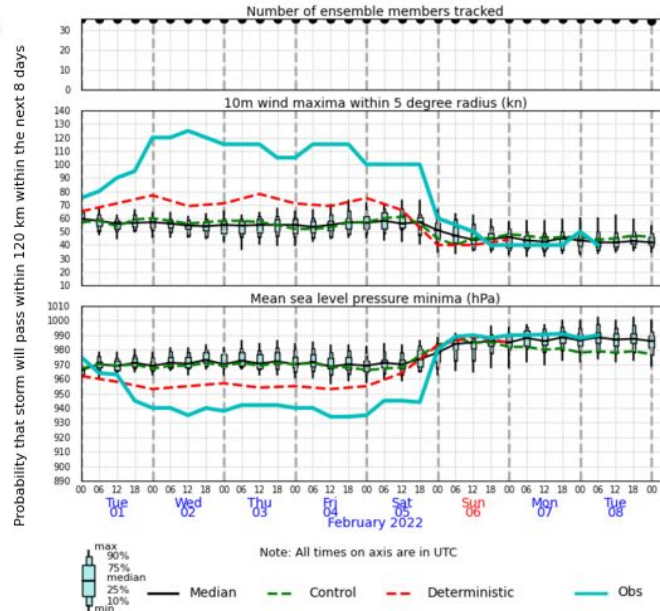


Extremes - Tropical Cyclones

MOGREPS-G: Forecast TC track probability for BATSIRAI from 00 UTC Tue 01 Feb 2022



MOGREPS-G ensemble: TC-following meteorogram BATSIRAI (16.65 62.5E) from 00 UTC Tue 01 February 2022



- Batsarai Feb 2022
- 20 km ensemble with 10km deterministic
- Deterministic performs better than ensemble
- Need higher resolution for extreme cases

Conclusion

- There is a **diminishing return of skill** when you increase ensemble size
- Using less members in R&D does not greatly affect skill
 - 19 vs 44 members \square **~2% difference in CRPS**
- **Resolution is important** for forecast skill
- Computationally, a many member low-res ensemble may be preferable
- However, the skill of surface processes, sub-grid processes and extreme events benefit from high resolution

Questions?

